Estimacion y Propiedades del Modelo Lineal

Econometria IAE

Ricardo Pasquini

Octubre 2023

Modelo Lineal

- Anteriormente introducimos al modelo lineal para caracterizar la relación entre la variable a explicar ($\mathbf{y}$) y el vector de variables explicativas ($\mathbf{x}$).

$$\mathbf{y} = \mathbf{x}'\beta + e$$

- Además introdujimos la existencia de una mejor proyección lineal, que era estimada a partir del método de OLS:

$$\hat{\beta} = \text{argmin}_{\beta} \sum_{i} e_i^2 = \text{argmin}_{\beta} \sum_{i} (y - x'\beta)^2$$

Plan

- Predicción y Bondad de Ajuste
- *Varianza* de nuestro modelo lineal. Este concepto resultará crítico para la formulación de tests de hipótesis.

Predicción y Residuo estimado

- Predicción

$$\hat{\mathbf{y}} = \mathbf{x}'\hat{\beta}$$

- Residuo estimado

$$\hat{e} = \mathbf{y} - \hat{\mathbf{y}}$$

Bondad de Ajuste

1. **Error Cuadrático Medio (M.S.E.)**

$$\text{MSE} = \frac{1}{n} \sum_i \hat{e}_i{}^2$$

- Nota: Veremos que un estimador insesgado de $E[e^2] = \frac{1}{n-k} \sum_i \hat{e}_i{}^2$

Bondad de Ajuste

2. R^2

- Partiendo de la relación:

$$Y_i = \hat{Y}_i + e_i$$

* Se puede mostrar que se cumple:

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum e_i^2$$

* Es decir:

$$\text{SCTotales} = \text{SCExplicado} + \text{SCResiduos}$$

Bondad de Ajuste

Por lo tanto, una medida de bondad de ajuste es:

$$R^2 = \frac{\text{SCExplicado}}{\text{SCTotales}} = 1 - \frac{\text{SCResiduos}}{\text{SCTotales}}$$

Bondad de Ajuste

Por definición:

$$R^2 = 1 - \frac{\sum \hat{e_i^2}}{\sum (y_i - \bar{y}_i)^2}$$

Pero cuando incluimos muchas variables esta medida pierde sentido.

Bondad de Ajuste R^2 ajustado

La arreglamos ajustando por grados de libertad:

$$\bar{R}^2 = 1 - \frac{\sum \hat{e_i^2}/(n-k)}{\sum (y_i - \bar{y}_i)^2/(n-1)}$$

Modelo Lineal

- Anteriormente vimos que la mejor proyección lineal (en la población) es:

$$\beta = (E(xx')^{-1})E(xy)$$

- Podemos derivar el estimador OLS reemplazando por los estimadores de los valores esperados respectivos

$$\hat{\beta} = ((\frac{\sum x_i x_i'}{n})^{-1})(\frac{\sum x_i y_i}{n}) = (\sum x_i x_i')^{-1}(\sum x_i y_i)$$

- Este método se conoce como *plug-in*.
- Lo que hicimos fue estimar una función de la media poblacional $g(\mu)$ usando la media de la funciones $g(\hat{\mu}) = \frac{\sum g(y_i)}{n}$, una extensión de la Ley Grandes Números.

El estimador de OLS es insesgado

- En promedio el estimador del OLS obtiene el valor de β correcto.

$$E[\hat{\beta}] = \beta$$

- La demostración usa las propiedades de la esperanza:

$$E[\hat{\beta}] = E[(X'X)^{-1}X'Y] = E[(X'X)^{-1}X'(X\beta + e)] = \beta + E[X'e] = \beta$$

- En general no todos los estimadores que encontrarán serán insesgados, y esta es una propiedad deseable.

Varianza poblacional

$$var(X) = E[(X - E[X])^2]$$

- **Ejemplo:** Habíamos visto que si x es el valor que surge de la tirada de un dado, entonces $E[x] = 3.5$
- Varianza de x:

$$var(x) = E[(x - E[x])^2] = E[(x - 3.5)^2] =$$
$$\frac{1}{6}(1 - 3.5)^2 + \frac{1}{6}(2 - 3.5)^2 + ... + \frac{1}{6}(6 - 3.5)^2 = 2.916$$

Varianza de $\hat{\beta}$

- La varianza del estimador, en general, es

$$V_{\hat{\beta}} = \mathrm{var}(\hat{\beta}|X) = (X'X)^{-1}(X'DX)(X'X)^{-1}$$

donde

- X es una matriz formada por las variables explicativas, y
- D es la matriz que tiene $E(e_i^2|x_i) = \sigma_i^2$ en la diagonal y $E(e_i e_j|X)$ fuera de la diagonal.
- Por el momento σ_i^2 es desconocida

Propiedades del Estimador de OLS

- Notar que si las observaciones son *independendientes* y estamos en el caso de *homocedasticidad*

$$V_{\hat{\beta}} = \mathrm{var}(\hat{\beta}|X) = (X'X)^{-1}\sigma^2$$

  - Donde σ^2 es la varianza (asumida constante) del error. Por esta razón, en general reducir el error del modelo (lo no-explicado) ayudará a mejorar la precisión de la estimación de los coeficientes.

Homoscedasticidad vs Heteroscedasticidad

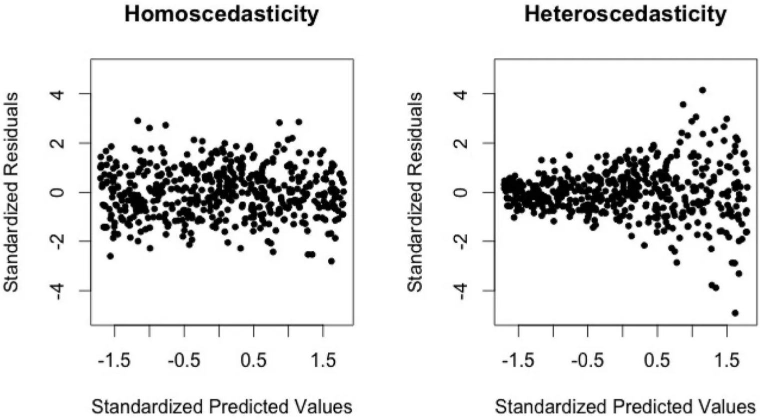

Figure 1: Analisis grafico

Varianza del error OLS

- La varianza del error es

$$\text{Var}(e|X) = E(ee'|X) \overset{\text{def}}{=} D$$

- Donde D es la matriz que tiene a $E(e_i^2|x_i) = \sigma_i^2$ en la diagonal, y $E[e_i e_j|X]$=0 fuera de la misma (debido a la independencia de las observaciones). También se puede denotar $D = diag(\sigma_1, \sigma_2, ..., \sigma_n)$
- Para el caso homocedástico:

$$Var(e|X) = I_n \sigma^2$$

Varianza del estimador de OLS

- En general la varianza de $\hat{\beta}$ es

$$V_{\hat{\beta}} = \mathrm{var}(\hat{\beta}|X) = (X'X)^{-1}(X'DX)(X'X)^{-1}$$

donde $D = E(\mathbf{ee}'|X)$ o dicho de otra forma la matriz que tiene $E(e_i^2|x_i) = \sigma_i^2$ en la diagonal y $E(e_i e_j|X)$ fuera de la diagonal.

$$(X'DX) = \sum_i^n x_i x_i' \sigma_i^2$$

Estimador de la Varianza del Error

- Recordemos que la varianza del error es equivalente al valor esperado del error al cuadrado. Queremos estimar $E(e_i^2|x_i)$. También podemos usar el estimador de momentos:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^2$$

- Resulta que el estimador es sesgado cuando cuando tengo pocas observaciones. Intuitivamente, para estimar el error se deben estimar previamente otros k parámetros, por lo que se pierden grados de libertad. Se ajusta de la siguiente forma:

$$s^2 = \left(\frac{1}{n-k}\right)\sum_{i=1}^{n}\hat{e}_i^2$$

Estimación de Var($\hat{\beta}$) (bajo homocedasticidad)

- Bajo homocedasticidad, la matriz de covarianza toma la forma

$$\text{Var}(\hat{\beta}) = (X'X)^{-1}\sigma^2$$

  - Podemos utilizar el estimador insesgado de σ^2 : s^2

$$\hat{\text{V}}\text{ar}(\hat{\beta}) = (X'X)^{-1}s^2$$

  - Este suele ser el estimador de default en muchos programas, incluso en STATA. Sin embargo hay que recordar que solo aplica bajo homocedasticidad (osea, condiciones bastante excepcionales).

Estimación de Var($\hat{\beta}$) (bajo heterocedasticidad)

- Bajo heterocedasticidad, la matriz de covarianza toma la forma

$$V_{\hat{\beta}} = \mathrm{var}(\hat{\beta}|X) = (X'X)^{-1}(X'DX)(X'X)^{-1}$$

  - O también

$$V_{\hat{\beta}} = \mathrm{var}(\hat{\beta}|X) = (X'X)^{-1}\sum_{i}^{n} x_i x_i' e_i^2 (X'X)^{-1}$$

Estimación de Var($\hat{\beta}$) (bajo heterocedasticidad)

- Es fácil mostrar que si conocieramos los errores (quod-non), podríamos construir directamente un estimador insesgado.
- Un estimador *factible* podría reemplazar los errores estimados que surgen de OLS, simplemente $\hat{e}_i$:

$$V_{\hat{\beta}}^{White} = (X'X)^{-1}(\sum_{i}^{n} x_i x_i' E(\hat{e}_i^2|X))(X'X)^{-1}$$

- El estimador de *White* es muy utilizado y es la opción "robust" en Stata.

Estimación de Var($\hat{\beta}$) (bajo heterocedasticidad)

- Como sabíamos que el estimador de la varianza del error es sesgado, una variante podría ser ajustar este estimador multiplicando por $n/(n-k)$

Scaled-White

$$V_{\hat{\beta}}^{\text{Scaled-White}} = \frac{n}{n-k}(X'X)^{-1}(\sum_{i}^{n} x_i x_i' E(\hat{e}_i^2|X))(X'X)^{-1}$$

- Otras especificaciones usualmente utilizadas incorporan los principios de utilizar estimaciones del error mediante predicción (forecast error) , o residuos estandarizados. (Ver opciones hc2 y hc3 en Stata).

Mínimos Cuadrados Generalizados

- Consideremos ahora una situación donde los errores, en general, podrían ser heterocedásticos, o incluso estar correlacionados (i.e., no ser independientes).

$$
\begin{aligned}
y &= X\beta + e \\
E(e|X) &= 0 \\
\text{Var}(e|X) &= \Omega
\end{aligned}
$$

Mínimos Cuadrados Generalizados

- Bajo estos supuestos se puede mostrar que las propiedades de OLS en este caso son

$$E(\hat{\beta}|X) = \beta$$
$$\mathrm{Var}(\beta|X) = (X'X)^{-1}(X'\Omega X)(X'X)^{-1}$$

- Usando el teorema de Gauss-Markov se puede ver que OLS es ineficiente, pues la varianza no es la mínima posible (que para este caso sería $(X'\Omega^{-1}X)^{-1}$)

Mínimos Cuadrados Generalizados

- Suponiendo que conocemos Ω en una cierta escala (i.e., suponemos que $\Omega = c^2 * \Sigma$ con Σ conocido.) Es posible derivar un estimador alternativo que logra la mínima varianza.

- El procedimiento consiste en premultiplicar el modelo por $\Sigma^{-1/2}$, y obtener:

$$\tilde{y} = \tilde{X}\beta + \tilde{e}$$

donde $\tilde{y} = \Sigma^{-1/2}y$, etc.

- Se puede mostrar fácilmente que esto resulta en:

$$\tilde{\beta}_{GLS} = (X'\Sigma^{-1/2}X)^{-1}(X'\Sigma^{-1/2}y)$$

- Este estimador redunda en la menor varianza, pero como no conocemos Σ, no es factible. Pero en la práctica se utilizan varias formas de aproximar Σ.

Test de Hipótesis

- Un test de hipótesis de especial interés será convalidar si (en la población) $\beta_j = 0$.
- Lo formalizamos con el test

$$H_0 : \beta_j = 0$$
$$H_1 : \beta_j \neq 0$$

- Podemos testearlo con:

$$t = \frac{b_j}{\text{std.error}(\beta)} = \frac{b_j}{\sqrt{\hat{V}ar(\beta)}} \sim T_{n-k-1}$$

Modelo de Regresión *Normal*

- Incorporando el supuesto de que los errores se distribuyen normalmente

$$\frac{\widehat{\beta}_j - \beta_j}{s(\hat{\beta}_j)} = \frac{\widehat{\beta}_j - \beta_j}{s\sqrt{\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right]_{jj}}} \sim \frac{\mathrm{N}\left(0, \sigma^2\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right]_{jj}\right)}{\sqrt{\frac{\sigma^2}{n-k}\chi^2_{n-k}}\sqrt{\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right]_{jj}}} = \frac{\mathrm{N}\left(0,1\right)}{\sqrt{\frac{\chi^2_{n-k}}{n-k}}} \sim t_{n-k}$$

a t distribution with $n-k$ degrees of freedom.

Figure 2: Test

Distribución T

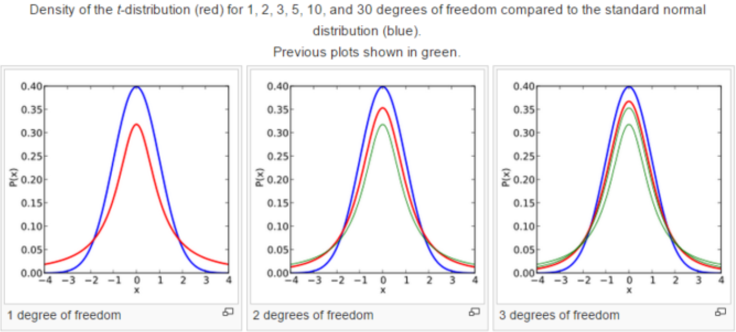

Figure 3: T-distribution

Conclusiones

- La variabilidad de los coeficientes estimados depende de la variabilidad del error: intentar bajar el error del modelo (por ejemplo, modificando la forma funcional o agregando factores explicativos) implicará una reducción de la variabilidad de los estimadores.
  - La variabilidad también depende (de una forma compleja) de las variables explicativas incorporadas al modelo y de la relación entre las mismas. Por el momento no podemos decir mucho más, pero volveremos sobre esto.

Conclusiones

- El Test de Hipótesis permite sacar una conclusión al respecto de la *significancia estadística* de un coeficiente. La significancia estadística es un criterio *relativo*: para sacar una conclusión respecto a una hipótesis se tiene en cuenta cuán grande o pequeño es un coeficiente en relación a la variabilidad de su medición.
  - Esta idea queda reflejada en el hecho de que la decisión sobre el test se basa en el valor del estadístico "T". Nótese que "T" es un ratio y como tal puede interpretarse como el valor del coeficiente medido en unidades de su variabilidad.
  - Por esta razón el test no dice nada sobre si una magnitud estimada es económicamente importante o no, y viceversa: una magnitud económicamente importante no es suficiente para afirmar algo al respecto de la significancia estadística.

Conclusiones

- El análisis del error (y sus supuestos) es importante en cuanto de ello depende la elección del estimador de varianza apropiado y la validez de los supuestos del test de hipótesis:
  - Ante la presencia de heterocedasticidad en los errores, optar por un estimador *robusto* de la varianza. Esto tiene el costo de mayor varianza que un estimador no robusto.
  - Identificar la no-normalidad en los errores es problemático especialmente cuando se cuenta con muestras pequeñas. Se recomienda buscar transformar el modelo en busca de mayor normalidad. Veremos más sobre esto en la próxima. En muestras grande el problema es menor.

Referencias

- Hansen, B. 2018, Econometrics